АНАЛИЗ СТРУКТУРЫ *D*-ОПТИМАЛЬНЫХ ПЛАНОВ ЭКСПЕРИМЕНТОВ С НЕРАВНОТОЧНЫМИ НАБЛЮДЕНИЯМИ

В. П. Кирлица

Белорусский государственный университет
Минск, Беларусь
E-mail: Kirlitsa@bsu.by

В статье приводится структура точных *D*-оптимальных планов экспериментов для линии регрессии с некоррелированными, неравноточными наблюдениями, когда дисперсии наблюдений зависят от точек наблюдения, и для случая, когда дисперсии постоянны, но отличаются друг от друга. Для дисперсий, зависящих от точек проведения эксперимента, выделен класс функций, описывающих изменение дисперсий, для которого *D*-оптимальный план экспериментов будет таким же, как и для равноточных наблюдений. Если же дисперсии наблюдений постоянны в каждом эксперименте, то *D*-оптимальный план экспериментов может существенно отличаться от плана для равноточных наблюдений.

Ключевые слова: некоррелированные, неравноточные наблюдения, точный *D*-оптимальный план экспериментов.

ВВЕДЕНИЕ

Рассмотрим модель парной регрессии с неравноточными наблюдениями

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i,\ i = 1,\ldots,n,\ n \geq 2, \tag{1}$$

где y_i – наблюдаемые значения; θ_0, θ_1 – неизвестные параметры; x_i – контролируемые переменные из интервала [–1, 1]; ε_i – некоррелированные ошибки наблюдений с нулевым средним и дисперсией $\sigma_i^2 > 0$. Модель наблюдений (1) означает, что в i-ом эксперименте наблюдения проводятся равноточно с одной и той же дисперсией σ_i^2. Однако, при переходе от одного эксперимента к другому точность измерений может изменяться. Результаты одного эксперимента не зависят от результатов других экспериментов.

Цель данной статьи – построить точные (n – задано) D-оптимальные планы экспериментов для оценивания неизвестных параметров θ_0, θ_1 модели (1) некоррелированных, неравноточных наблюдений. Точные D-оптимальные планы экспериментов для модели некоррелированных, равноточных наблюдений (1) построены в [1].

В работе автора статьи [2] исследовалась структура точных D-оптимальных планов экспериментов для модели наблюдений (1) в случае некоррелированных, неравноточных наблюдений, дисперсии которых зависят от точек x_i проведения эксперимента, $D\{\varepsilon_i\} = d(x_i) > 0$ для каждой реализации $x_i \in [-1,1]$. Класс функций $d(x)$ определяется следующим образом:

$$d(x) \ge \frac{1}{4}\{[d_1 + d_2]x^2 + 2[d_2 - d_1]x + d_1 + d_2\}, d_1 = d(-1) > 0, d_2 = d(1) > 0,\ x \in [-1, 1]. \quad (2)$$

Класс функций (2) определяет дисперсии $d(x)$, которые лежат не ниже минорантной параболы

$$\frac{1}{4}\{[d_1 + d_2]x^2 + 2[d_2 - d_1]x + d_1 + d_2\}. \quad (3)$$

Классу функций $d(x)$ удовлетворяют равноточные наблюдения ($d(x) = d = const$), линейное изменение дисперсии наблюдений ($d(x) = a + bx$, $a > 0$, $|b| < a$), а также все вогнутые функции, неотрицательные на интервале [–1, 1].

В [2] установлено, что структура точных D-оптимальных планов экспериментов для модели наблюдений (1) с неравноточными наблюдениями, дисперсия которых $d(x)$ удовлетворяет (2), зависит от того, четно или нет число наблюдения n. Точный D-оптимальный план (n фиксировано) определяется следующим образом: для четного числа наблюдений $n = 2s$

$$\varepsilon_{2s}^0 = \begin{Bmatrix} -1; & 1 \\ s; & s \end{Bmatrix}, \quad (4)$$

для нечетного числа наблюдений $n = 2s + 1$

$$\varepsilon_{2s+1}^0 = \begin{Bmatrix} -1; & 1 \\ s; & s+1 \end{Bmatrix}, \varepsilon_{2s+1}^0 = \begin{Bmatrix} -1; & 1 \\ s+1; & s \end{Bmatrix}. \quad (5)$$

НЕКОРРЕЛИРОВАННЫЕ, НЕРАВНОТОЧНЫЕ НАБЛЮДЕНИЯ ДЛЯ КАЖДОГО ЭКСПЕРИМЕНТА

Перейдем к более детальному исследованию структуры точных D-оптимальных планов экспериментов для модели (1) некоррелированных наблюдений с дисперсиями $D\{y_i\} = \sigma_i^2 > 0$, $i = \overline{1, n}$.

Лемма. Для модели некоррелированных наблюдений (1) с дисперсиями $D\{y_i\} = \sigma_i^2 > 0$, $i = \overline{1, n}$, точки спектра x_i^0 точного D-оптимального плана экспериментов лежат на концах интервала [–1, 1], т.е. $x_i^0 = \pm 1$, $i = \overline{1, n}$.

Доказательство. В матричной форме записи модель (1) представима в виде [3]:

$$Y = X\theta + \varepsilon,\ E\varepsilon = O_n,\ \ \Sigma_\varepsilon = \Sigma_0, \quad (6)$$

где X – матрица плана экспериментов; $\theta' = (\theta_0, \theta_1)$ – вектор неизвестных параметров; ε – n-вектор ошибок наблюдений; $O_n = (0,\ldots,0)$ – n-вектор-столбец, состоящий из одних нулей; Σ_ε – ковариационная матрица ошибок наблюдений; Σ_0 – симметричная, диагональная матрица с i-тыми диагональными элементами $\sigma_i^2 > 0$, $i = \overline{1, n}$.

Точность оценки, полученной по обобщенному методу наименьших квадратов, определяется ковариационной матрицей [3]:

$$\Sigma_{\hat\theta} = (X'\Sigma_0^{-1}X)^{-1}.$$

Точный D-оптимальный план экспериментов соответствует плану экспериментов X, для которого определитель

$$\left|X'\Sigma_0^{-1}X\right| = \sum_1^n \frac{1}{\sigma_i^2}\sum_1^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_1^n \frac{x_i}{\sigma_i^2}\right)^2 \tag{7}$$

достигает своего максимального значения на множестве $x_i \in [-1,\ 1]$, $i = \overline{1,n}$. Покажем, что максимальное значение определителя (4) достигается при $x_i = \pm 1$, $i = \overline{1,n}$.

Сделаем некоторую точку x_i, $i = \overline{1,n}$, "плавающей" на интервале [–1, 1]. Пусть, например, не ограничивая общности рассуждений, $x_1 = x \in [-1,\ 1]$, а остальные значения x_i, $i = \overline{2,n}$ фиксированы. Тогда

$$f(x) = \sum_1^n \frac{1}{\sigma_i^2}\left(\frac{x^2}{\sigma_1^2} + \sum_2^n \frac{x_i^2}{\sigma_i^2}\right) - \left(\frac{x}{\sigma_1^2} + \sum_2^n \frac{x_i}{\sigma_i^2}\right) = \left(\frac{1}{\sigma_1^2}\sum_2^n \frac{1}{\sigma_i^2}\right)x^2 -$$

$$2\left(\sum_2^n \frac{x_i}{(\sigma_1\sigma_i)^2}\right)x + \sum_1^n \frac{1}{\sigma_i^2}\sum_2^n \frac{x_i^2}{\sigma_i^2} - \left(\sum_2^n \frac{x_i}{\sigma_i^2}\right)^2. \tag{8}$$

Из (8) следует, что $f(x)$ – выпуклая функция, какие бы ни были значения x_i, $i = \overline{2,n}$. Максимум этой функции достигается при $x = \pm 1$. Лемма доказана.

Опираясь на лемму, докажем следующую теорему.

Теорема. Точки спектра x_i^0, $i = \overline{1,n}$, точного *D*-оптимального плана экспериментов с неравноточными, некоррелированными наблюдениями (1) лежат на концах интервала [–1, 1], $x_i^0 = \pm 1$, $i = \overline{1,n}$. И это такие комбинации точек, для которых абсолютная величина

$$\left|\sum_1^n \frac{x_i^0}{\sigma_i^2}\right| \tag{9}$$

принимает минимальное значение.

Доказательство. Для точек спектра точного *D*-оптимального плана $x_i^0 = \pm 1$ определитель (7) принимает значение

$$\left(\sum_1^n \frac{1}{\sigma_i^2}\right)^2 - \left(\sum_1^n \frac{x_i^0}{\sigma_i^2}\right)^2. \tag{10}$$

Максимальное значение (10) достигается, когда модуль суммы в (9) принимает минимальное значение. Теорема доказана.

Из теоремы следует, что задача построения точных *D*-оптимальных планов экспериментов для модели (1) с некоррелированными, неравноточными наблюдениями сводится к оптимизационной задаче целочисленного программирования, т.е. к минимизации выражения (9) на множестве $x_i^0 = \pm 1$, $i = \overline{1,n}$. В общем случае, при численном построении *D*-оптимальных планов экспериментов, надо произвести 2^n вычислений выражения (9). Рассмотрим теперь некоторые частные случаи, в которых задача построения *D*-оптимальных планов экспериментов значительно упрощается.

Для равноточных наблюдений, $\sigma_i^2 = \sigma^2$, $i = \overline{1,n}$, нет необходимости организовывать оптимизационную процедуру. Результат сформулируем в виде следствия из теоремы.

Следствие. Для равноточных ($\sigma_i^2 = \sigma^2$, $i = \overline{1,n}$), некоррелированных наблюдений (1) точный *D*-оптимальный план экспериментов зависит от четности проведенных экспериментов *n*, не зависит от значения σ^2 и определяется соотношениями (4), (5).

Рассмотрим теперь другой частный случай некоррелированных, неравноточных наблюдений, когда серия из n независимых наблюдений подразделяется на две независимых серий наблюдений: $y_1, \ldots, y_{n_1}; \overline{y_1}, \ldots, \overline{y_{n_2}}$, $n = n_1 + n_2$. Наблюдения $y_1, \ldots, y_{n_1}$ проводятся в точках $x_1, \ldots, x_{n_1}$ с равными дисперсиями σ_1^2, а наблюдения $\overline{y_1}, \ldots, \overline{y_{n_2}}$ проводятся в точках $\overline{x_1}, \ldots, \overline{x_{n_2}}$ с равными дисперсиями σ_2^2.

В этом случае проблема построения точных D-оптимальных планов экспериментов для модели наблюдений (1) сводится к минимизации по $x_i^0 = \pm 1$, $i = \overline{1, n_1}$; $\overline{x}_i^0 = \pm 1$, $i = \overline{1, n_2}$ выражения

$$\left|\sum_1^{n_1} \frac{x_i^0}{\sigma_1^2} + \sum_1^{n_2} \frac{\overline{x}_i^0}{\sigma_2^0}\right| = \frac{1}{\sigma_2^2}\left|\overline{\sigma}\sum_1^{n_1} x_i^0 + \sum_1^{n_2} \overline{x}_i^0\right|, \tag{11}$$

где $\overline{\sigma} = \dfrac{\sigma_2^2}{\sigma_1^2}$. Обозначим через k число x_i^0, принимающих значение –1, а через m – число $\overline{x}_i^0$, принимающих значение –1, $0 \le k \le n_1$, $0 \le m \le n_2$. Тогда

$$\min\left|\overline{\sigma}\sum_1^{n_1} x_i^0 + \sum_1^{n_2} \overline{x}_i^0\right| = \min\left|\overline{\sigma}(n_1 - 2k) + n_2 - 2m\right|, \tag{12}$$

где минимум по k и m берется на множестве: $0 \le k \le n_1$, $0 \le m \le n_2$. Решение оптимизационной задачи (12) и определяет структуру точных D-оптимальных планов экспериментов для неравноточных наблюдений, разбитых на две группы равноточных наблюдений. В данном случае необходимо произвести $(n_1+1)(n_2+1)$ вычислений выражения, стоящего под знаком модуля в правой части (12), что значительно меньше чем 2^n, $n = n_1 + n_2$ вычислений. Однако, в этом случае надо знать отношение $\overline{\sigma}$ дисперсий σ_2^2 и σ_1^2.

Важно отметить, что, несмотря на то, что наблюдения не коррелированны в первой группе значений $y_1, \ldots, y_{n_1}$ и во второй группе значений $\overline{y_1}, \ldots, \overline{y_{n_2}}$, планировать D-оптимальным образом эксперименты внутри каждой группы по отдельности нельзя. Убедиться в этом можно на следующем примере.

Пусть проводится 5 наблюдений в первой группе экспериментов с одной и той же дисперсией $\sigma_1^2 = 1$ и 4 наблюдения во второй группе экспериментов с дисперсией $\sigma_2^2 = 2$ для всех наблюдений из второй группы.

Решение оптимизационной задачи (12) дает следующий результат. Точных D-оптимальных планов экспериментов два. Оптимальные значения k^0 и m^0 следующие: $k^0 = 2$, $m^0 = 3$ либо $k^0 = 3$, $m^0 = 1$. Для этих значений минимум в (12) равен нулю.

В первом D-оптимальном плане нужно провести 2 наблюдения в точке –1 и 3 наблюдения в точке 1 для первой группы экспериментов. Для второй группы экспериментов распределение наблюдений следующее. Три наблюдения проводятся в точке –1, а четвертое наблюдение – в точке 1.

Во втором D-оптимальном плане нужно провести 3 наблюдения в точке –1 и 2 наблюдения в точке 1 для первой группы экспериментов. Для второй группы экспериментов распределение наблюдений следующее. Одно наблюдение проводится в точке –1, а три наблюдения – в точке 1.

Если бы ориентироваться на *D*-оптимальную организацию измерений в каждой группе по отдельности, то оптимальные значения k^0 и m^0 были бы равны: $k^0 = 2$, $m^0 = 2$ либо $k^0 = 3$, $m^0 = 2$. Но при этих значениях k^0 и m^0 имеем:

$$\left|14 - 4k^0 - 2m^0\right| = 2 \neq 0.$$

Следовательно, предложенный вариант плана экспериментов не является *D*-оптимальным.

ЛИТЕРАТУРА

1. *Федоров, В.* Теория оптимального эксперимента / В. Федоров. М. : Наука, 1971. 312 с. (С. 66).
2. *Кирлица, В. П.* Структура точных *D*-оптимальных планов экспериментов для линии регрессии с неравноточными наблюдениями / В. П. Кирлица // Теория вероятностей, случайные процессы, математическая статистика и их приложения. Мн. : РИВШ, 2014. С. 67–70.
3. *Айвазян, С.* Эконометрика / С. Айвазян, С. Иванова. М. : Маркет ДС, 2007. 104 с. (С.30–32).